



## 关系抽取技术研究综述

黄勋, 游宏梁, 于洋

中国国防科技信息中心 北京 100142

Huang Xun, You Hongliang, Yu Yang

China Defense Science &amp; Technology Information Center, Beijing 100142, China

- 摘要
- 参考文献
- 相关文章

Download: PDF (6206KB) [HTML](#) (1KB) Export: BibTeX or EndNote (RIS) Supporting Info

**摘要** 对关系抽取技术研究概况进行总结。在回顾关系抽取发展历史的基础上, 将关系抽取研究划分为两个阶段: 面向特定领域的关系抽取研究和面向开放互联网文本的关系抽取研究。在分析相关文献的基础上, 总结出两个研究阶段的技术路线: 面向特定领域的关系抽取技术以基于标注语料的机器学习方法为主; 面向开放互联网文本的关系抽取则根据不同任务需要, 采取基于启发式规则的方法或者基于背景知识库实例的机器学习方法。

关键词: [关系抽取](#) [信息抽取](#) [机器学习](#)

**Abstract:** The paper summarizes the research of relation extraction technology. It firstly gives a brief overview of relation extraction, and divides the research into two phases: the relation extraction in specific domains and the relation extraction of Web text. Then, analyzes the major methodologies of the two phases: the relation extraction in specific domains mainly uses machine learning methods with annotated corpora, while the relation extraction of Web text uses rule-based methods or distant supervision methods according to different demands.

Keywords: [Relation extraction](#), [Information extraction](#), [Machine learning](#)

收稿日期: 2013-07-12;

引用本文:

黄勋, 游宏梁, 于洋. 关系抽取技术研究综述[J]. 现代图书情报技术, 2013, V29(11): 30-39

Huang Xun, You Hongliang, Yu Yang. A Review of Relation Extraction[J], 2013, V29(11): 30-39

链接本文:

<http://www.infotech.ac.cn/CN/> 或 <http://www.infotech.ac.cn/CN/Y2013/V29/I11/30>






- [1] Message Understanding Conference[EB/OL].[2013-06-24]. [http://en.wikipedia.org/wiki/Message\\_Understanding\\_Conference](http://en.wikipedia.org/wiki/Message_Understanding_Conference).
- [2] MUC-7 Information Extraction Task Definition[EB/OL].[2013-06-24]. [http://www.itl.nist.gov/iaui/894.02/related\\_projects/muc/proceedings/ie\\_task.html](http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/ie_task.html).
- [3] Automatic Content Extraction(ACE) Evaluation[EB/OL].[2013-06-24]. <http://www.itl.nist.gov/iad/mig//tests/ace/>.
- [4] The ACE 2007(ACE2007) Evaluation Plan[EB/OL].[2013-06-24]. <http://www.itl.nist.gov/iad/mig//tests/ace/ace07/doc/ace07-evalplan.v1.3a.pdf>.
- [5] Knowledge Base Population (KBP) 2013 [EB/OL].[2013-06-24]. <http://www.nist.gov/tac/2013/KBP/>.
- [6] Mintz M, Bills S, Snow R, et al. Distant Supervision for Relation Extraction Without Labeled Data[C].In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Association for Computational Linguistics, 2009: 1003- 1011.
- [7] Banko M. Open Information Extraction for the Web[D]. University of Washington, 2009.
- [8] Automatic Content Extraction 2008 Evaluation Plan (ACE08)[EB/OL].[2013-08-24]. <http://www.itl.nist.gov/iad/mig/tests/ace/2008/doc/ace08-evalplan.v1.2d.pdf>.
- [9] Aone C, Halverson L, Hampton T, et al. SRA: Description of the IE2 System Used for MUC-7[C].In: *Proceedings of the 7th Message Understanding Conference (MUC-7)*. 1998.
- [10] Fukumoto F, Shimohata M, Masui F, et al. Oki Electric Industry: Description of the Oki System as Used for MET-2[C].In: *Proceedings of the*



## Service


- ▶ 把本文推荐给朋友
- ▶ 加入我的书架
- ▶ 加入引用管理器
- ▶ Email Alert
- ▶ RSS

## 作者相关文章

- ▶ 黄勋
- ▶ 游宏梁
- ▶ 于洋

- [11] Humphreys K, Gaizauskas R, Azzam S, et al. University of Sheffield: Description of the LaSIE-II System as Used for MUC-7[C].In: *Proceedings of the 7th Message Understanding Conference (MUC-7)*. 1998.
- [12] Miller S, Fox H, Ramshaw L, et al. A Novel Use of Statistical Parsing to Extract Information from Text[C].In: *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference*. Association for Computational Linguistics, 2000: 226-233.
- [13] Kambhatla N. Combining Lexical, Syntactic, and Semantic Features with Maximum Entropy Models for Extracting Relations[C].In: *Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2004.
- [14] 车万翔, 刘挺, 李生. 实体关系自动抽取[J]. 中文信息学报, 2005,19(2): 1-6. (Che Wanxiang, Liu Ting, Li Sheng. Automatic Entity Relation Extraction [J]. *Journal of Chinese Information Processing*, 2005,19(2): 1-6.) 
- [15] Zhou G D, Su J, Zhang J, et al. Exploring Various Knowledge in Relation Extraction[C].In: *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics,2005: 427-434.
- [16] Jiang J, Zhai C X. A Systematic Exploration of the Feature Space for Relation Extraction[C].In: *Proceedings of Human Language Technologies: The Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT'07)*. 2007:113-120.
- [17] 董静, 孙乐, 冯元勇, 等. 中文实体关系抽取中的特征选择研究[J]. 中文信息学报, 2007,21(4): 80-85. (Dong Jing, Sun Le, Feng Yuanyong, et al. Chinese Automatic Entity Relation Extraction[J]. *Journal of Chinese Information Processing*, 2007, 21(4): 80-85.) 
- [18] 陈宇, 郑德权, 赵铁军. 基于 Deep Belief Nets 的中文名实体关系抽取[J]. 软件学报, 2012,23(10): 2572-2585. (Chen Yu, Zheng Dequan, Zhao Tiejun. Chinese Relation Extraction Based on Deep Belief Nets[J]. *Journal of Software*, 2012,23(10): 2572-2585.)
- [19] Zelenko D, Aone C, Richardella A. Kernel Methods for Relation Extraction[J]. *The Journal of Machine Learning Research*, 2003,3:1083-1106.
- [20] Culotta A, Sorensen J. Dependency Tree Kernels for Relation Extraction[C].In: *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2004.
- [21] Bunescu R C, Mooney R J. A Shortest Path Dependency Kernel for Relation Extraction[C].In: *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics,2005: 724-731.
- [22] 黄瑞红, 孙乐, 冯元勇, 等. 基于核方法的中文实体关系抽取研究[J]. 中文信息学报, 2008, 22(5): 102-108. (Huang Ruihong, Sun Le, Feng Yuanyong, et al. A Study on Kernel-based Chinese Relation Extraction[J]. *Journal of Chinese Information Processing*, 2008, 22(5): 102-108.)
- [23] Zhang M, Zhang J, Su J, et al. A Composite Kernel to Extract Relations Between Entities with Both Flat and Structured Features[C].In: *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2006: 825-832.
- [24] Zhou G D, Zhang M, Ji D H, et al. Tree Kernel-based Relation Extraction with Context-Sensitive Structured Parse Tree Information[C].In: *Proceedings of the 2007 Joint Conference on Empirical methods in Natural Language Processing and Computational Natural Language Learning (EMNLP/CoNLL-2007)*. 2007: 728-736.
- [25] Qian L H, Zhou G D, Kong F, et al. Exploiting Constituent Dependencies for Tree Kernel-based Semantic Relation Extraction[C].In: *Proceedings of the 22nd International Conference on Computational Linguistics*. Association for Computational Linguistics, 2008:697-704.
- [26] 庄成龙, 钱龙华, 周国栋. 基于树核函数的实体语义关系抽取方法研究[J]. 中文信息学报, 2009, 23(1): 3-9. (Zhuang Chenglong, Qian Longhua, Zhou Guodong. Research on Tree Kernel-based Entity Semantic Relation Extraction[J]. *Journal of Chinese Information Processing*, 2009, 23(1): 3-9.)
- [27] 虞欢欢, 钱龙华, 周国栋, 等. 基于合一句法和实体语义树的中文语义关系抽取[J]. 中文信息学报, 2010, 24(5): 17-23. (Yu Huanhuan, Qian Longhua, Zhou Guodong, et al. Chinese Semantic Relation Extraction Based on Unified Syntactic and Entity Semantic Tree[J]. *Journal of Chinese Information Processing*, 2010,24(5): 17-23.)
- [28] 刘克彬, 李芳, 刘磊, 等. 基于核函数中文关系自动抽取系统的实现[J]. 计算机研究与发展, 2007, 44(8): 1406-1411. (Liu Kebin, Li Fang, Liu Lei, et al. Implementation of a Kernel-based Chinese Relation Extraction System[J]. *Journal of Computer Research and Development*, 2007, 44(8): 1406-1411.) 
- [29] Brin S. Extracting Patterns and Relations from the World Wide Web[C]. In: *Proceedings of International Workshop on the World Wide Web and Databases*. London, UK: Springer-Verlag, 1999: 172-183.
- [30] Agichtein E, Gravano L. Snowball: Extracting Relations from Large Plain-text Collections[C]. In: *Proceedings of the 5th ACM Conference on Digital Libraries*. ACM, 2000:85-94.
- [31] Etzioni O, Cafarella M, Downey D, et al. Unsupervised Named-entity Extraction from the Web: An Experimental Study[J]. *Artificial Intelligence*, 2005,165(1): 91-134. 
- [32] Rosenfeld B, Feldman R. URES: An Unsupervised Web Relation Extraction System[C].In: *Proceedings of the COLING/ACL on Main Conference Poster Sessions*. Association for Computational Linguistics,2006: 667-674.
- [33] Feldman R, Rosenfeld B. Boosting Unsupervised Relation Extraction by Using NER[C].In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2006: 473-481.
- [34] 何婷婷, 徐超, 李晶, 等. 基于种子自扩展的命名实体关系抽取方法[J]. 计算机工程, 2006, 32(21): 183-184,193. (He Tingting, Xu Chao, Li Jing, et al. Named Entity Relation Extraction Method Based on Seed Self-expansion[J]. *Computer Engineering*, 2006, 32(21): 183-184,193.) 

- [35] 李维刚, 刘挺, 李生. 基于网络挖掘的实体关系元组自动获取[J]. 电子学报, 2007, 35(11): 2111-2116. (Li Weigang, Liu Ting, Li Sheng. Automated Entity Relation Tuple Extraction Using Web Mining[J]. *Acta Electronica Sinica*, 2007, 35(11): 2111-2116.) 
- [36] Xu F Y, Uszkoreit H, Li H. A Seed-driven Bottom-up Machine Learning Framework for Extracting Relations of Various Complexity[C]. In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. 2007: 584-591.
- [37] Xu F Y. Bootstrapping Relation Extraction from Semantic Seeds[D]. Saarland University, 2008.
- [38] Xu F Y, Uszkoreit H, Li H, et al. Adaptation of Relation Extraction Rules to New Domains[C]. In: *Proceedings of the Poster Session of the 6th International Conference on Language Resources and Evaluation (LREC'08)*. 2008.
- [39] Xu F Y, Uszkoreit H, Krause S, et al. Boosting Relation Extraction with Limited Closed-world Knowledge[C]. In: *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 2010: 1354-1362.
- [40] Zhu J, Nie Z, Liu X J, et al. StatSnowball: A Statistical Approach to Extracting Entity Relationships[C]. In: *Proceedings of the 18th International Conference on World Wide Web*. ACM, 2009: 101-110.
- [41] Carlson A, Betteridge J, Wang R C, et al. Coupled Semi-supervised Learning for Information Extraction[C]. In: *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*. ACM, 2010: 101-110.
- [42] 陈锦秀, 姬东鸿. 基于图的非监督关系抽取[J]. 软件学报, 2008, 19(11): 2843-2852. (Chen Jinxiu, Ji Donghong. Graph-based Semi-Supervised Relation Extraction[J]. *Journal of Software*, 2008, 19(11): 2843-2852.) 
- [43] Curran J R, Murphy T, Scholz B. Minimising Semantic Drift with Mutual Exclusion Bootstrapping[C]. In: *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*. 2007: 172-180.
- [44] Hasegawa T, Sekine S, Grishman R. Discovering Relations Among Named Entities from Large Corpora[C]. In: *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2004.
- [45] Stevenson M. An Unsupervised WordNet-based Algorithm for Relation Extraction[C]. In: *Proceedings of the 4th International Conference on Language Resources and Evaluation Workshop "Beyond Named Entity: Semantic Labelling for NLP Tasks"*, Lisbon, Portugal. 2004.
- [46] Zhang M, Su J, Wang D, et al. Discovering Relations Between Named Entities from a Large Raw Corpus Using Tree Similarity-based Clustering [C]. In: *Proceedings of the 2nd International Joint Conference on Natural Language Processing (IJCNLP'05)*. Berlin, Heidelberg: Springer-Verlag, 2005: 378-389.
- [47] Rosenfeld B, Feldman R. Clustering for Unsupervised Relation Identification[C]. In: *Proceedings of the 16th ACM Conference on Information and Knowledge Management*. ACM, 2007: 411-418.
- [48] Davidov D, Rappoport A, Koppel M. Fully Unsupervised Discovery of Concept-specific Relationships by Web Mining[C]. In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. 2007: 232-239.
- [49] Yan Y, Okazaki N, Matsuo Y, et al. Unsupervised Relation Extraction by Mining Wikipedia Texts Using Information from the Web[C]. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Association for Computational Linguistics, 2009: 1021-1029.
- [50] Bollegala D T, Matsuo Y, Ishizuka M. Measuring the Similarity Between Implicit Semantic Relations from the Web[C]. In: *Proceedings of the 18th International Conference on World Wide Web*. ACM, 2009: 651-660.
- [51] Bollegala D T, Matsuo Y, Ishizuka M. Relational Duality: Unsupervised Extraction of Semantic Relations Between Entities on the Web[C]. In: *Proceedings of the 19th International Conference on World Wide Web*. ACM, 2010: 151-160.
- [52] Sekine S. On-Demand Information Extraction[C]. In: *Proceedings of the COLING/ACL on Main Conference Poster Sessions*. Association for Computational Linguistics, 2006: 731-738.
- [53] Shinyama Y, Sekine S. Preemptive Information Extraction Using Unrestricted Relation Discovery[C]. In: *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. Association for Computational Linguistics, 2006: 304-311.
- [54] Yates A, Cafarella M, Banko M, et al. TextRunner: Open Information Extraction on the Web[C]. In: *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. Association for Computational Linguistics, 2007: 25-26.
- [55] Wu F, Weld D. Autonomously Semantifying Wikipedia[C]. In: *Proceedings of the 16th ACM Conference on Information and Knowledge Management (CIKM'07)*, Lisbon, Portugal. 2007: 41-50.
- [56] Etzioni O, Fader A, Christensen J, et al. Open Information Extraction: The 2nd Generation[C]. In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*. AAAI Press, 2011: 3-10.
- [57] Schmitz M, Bart R, Soderland S, et al. Open Language Learning for Information Extraction[C]. In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, 2012: 523-534.
- [58] Ji H, Grishman R. Knowledge Base Population: Successful Approaches and Challenges[C]. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. 2011: 1148-1158.
- [59] Mintz M, Bills S, Snow R, et al. Distant Supervision for Relation Extraction Without Labeled Data[C]. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Association for Computational Linguistics, 2009: 1003-1011.

- [60] Krause S, Li H, Uszkoreit H, et al. Large-scale Learning of Relation-extraction Rules with Distant Supervision from the Web[C]. In: *Proceedings of the 11th International Semantic Web Conference*, Boston, MA, USA. Berlin, Heidelberg: Springer, 2012:263-278.
- [61] Akbik A, Visengeriyeva L, Herger P, et al. Unsupervised Discovery of Relations and Discriminative Extraction Patterns[C]. In: *Proceedings of COLING*.2012:17-32.
- [62] Nguyen T V T, Moschitti A. End-to-End Relation Extraction Using Distant Supervision from External Semantic Repositories[C]. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.2011:277-282.
- [63] Krishnamurthy R, Li Y Y, Raghavan S, et al. SystemT: A System for Declarative Information Extraction[J]. *ACM SIGMOD Record*,2009,37(4):7-13. 
- [64] Cunningham H, Maynard D, Tablan V, et al. JAPE: A Java Annotation Patterns Engine[OL]. [2013-06-24].<http://citeseerx.ist.psu.edu/viewdoc/download?jsessionid=B038120DE79C13635419187BFF58DFFF?doi=10.1.1.32.3929&rep=rep1&type=pdf>.
- [65] Li Y Y, Chiticariu L, Yang H, et al. WizIE: A Best Practices Guided Development Environment for Information Extraction[C].In: *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics,2012:109-114.
- [66] Chiticariu L, Krishnamurthy R, Li Y Y, et al. Domain Adaptation of Rule-based Annotators for Named-entity Recognition Tasks[C].In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics,2010:1002-1012.
- [67] Read the Web[EB/OL].[2013-07-07]. <http://rtw.ml.cmu.edu/rtw/>.
- [1] 翟东升, 张欣琦, 张杰, 康宁.分布式专利信息抽取系统设计与构建[J]. 现代图书情报技术, 2013,29(7/8): 114-121
- [2] 张晗, 刘双梅.中心度指标对语义述谓网络概念抽取的比较分析——以疾病治疗学研究为例[J]. 现代图书情报技术, 2013,(6): 30-35
- [3] 王秀艳, 崔雷.采用混合方法抽取生物医学实体间语义关系[J]. 现代图书情报技术, 2013,29(3): 77-82
- [4] 谷俊, 许鑫.中文专利中本体关系获取研究[J]. 现代图书情报技术, 2013,29(10): 73-78
- [5] 何琳, 何娟, 沈耕宇, 杨波, 黄水清.一种通过文本挖掘发现实时定量聚合酶链式反应实验内参基因的方法研究[J]. 现代图书情报技术, 2012,28(7/8): 109-114
- [6] 高强, 游宏梁.基于层叠模型的国防领域命名实体识别研究[J]. 现代图书情报技术, 2012,(11): 47-52
- [7] 王秀艳, 崔雷.应用关键词抽取生物医学实体间语义关系研究综述[J]. 现代图书情报技术, 2011,27(9): 21-27
- [8] 周虹, 张蓓, 姜爱蓉, 张成昱.馆藏书目信息自助短信推送服务的设计与实现[J]. 现代图书情报技术, 2011,27(7/8): 127-131
- [9] 王昊, 邓三鸿, 苏新宁.基于字序列标注的中文关键词抽取研究[J]. 现代图书情报技术, 2011,27(12): 39-45
- [10] 王志超, 翁楠, 王宇.基于主题句相似度的标题党新闻鉴别技术研究[J]. 现代图书情报技术, 2011,(11): 48-53
- [11] 逯万辉, 马建霞.基于条件随机场模型的复杂时间信息抽取研究[J]. 现代图书情报技术, 2011,27(10): 29-33
- [12] 孙镇 王惠临.命名实体识别研究进展综述[J]. 现代图书情报技术, 2010,26(6): 42-47
- [13] 聂卉 黄贵鹏.树编辑距离在Web信息抽取中的应用与实现\*[J]. 现代图书情报技术, 2010,26(5): 29-34
- [14] 詹佳佳.基于Web-Harvest的Web信息抽取系统的设计与应用[J]. 现代图书情报技术, 2010,26(3): 76-81
- [15] 武帅.基于条件随机场模型的事实信息抽取方法应用[J]. 现代图书情报技术, 2010,26(10): 59-64